Cell size effects in the molecular dynamics of the intrinsically disordered A peptide

Mehra, Rukmankesh; Kepp, Kasper Planeta

Published in:
Journal of Chemical Physics

Link to article, DOI:
10.1063/1.5115085

Publication date:
2019

Document Version
Peer reviewed version

Cell size effects in the molecular dynamics of the intrinsically disordered Aβ peptide

Rukmankesh Mehra and Kasper P. Kepp*

Technical University of Denmark, DTU Chemistry, Building 206, 2800 Kgs. Lyngby, Denmark.

* Correspondence to Kasper P. Kepp, e-mail: kpj@kemi.dtu.dk

ORCID of Authors

0000-0001-6010-1514, 0000-0002-6754-7348

ABSTRACT

Periodic molecular dynamics (MD) simulations of proteins may suffer from image interactions. Similarly, the hydrophobic effect required to keep a protein folded may not be enforced by small simulation cells. Accordingly, errors may arise both from the water concentration *per se* and the image interactions. Intrinsically disordered proteins are particularly sensitive, providing a worst-case estimate of the errors. Following this reasoning, we studied $A\beta_{40}$ (Aβ), a disordered peptide central to Alzheimer's disease, by 100 different simulations with variable cell size from very large (20 Å) to very small (3 Å). Even for this very disordered peptide, most properties are not cell-size dependent, justifying the common use of modest-sized (10 Å) cells for simulating proteins. The radius of gyration, secondary structure, intra-peptide, and peptide-water hydrogen bonds are similar relative to standard deviations at any cell size. However, hydrophobic surface area increases significantly in small cells (confidence 95%, two-tailed t-test), as does the standard deviation in exposure and backbone conformations (>40% and >27%). Similar results were obtained for the force fields OPLS3e, Ambersb99-ILDN, and Charmm22*. The similar prevalence of structures and α-β transitions in long and short simulations indicate small diffusion barriers, which we suggest is a defining hallmark of intrinsically disordered proteins. Whereas hydrophilic exposure dominates in large cells, hydrophobic exposure dominates in small cells, suggesting a weakening of the hydrophobic effect by image interactions and the few water layers available to keep the protein compact, with a critical limit of 2-3 water layers required to enforce the hydrophobic effect.

Keywords: Molecular dynamics, box size, image interactions, Aβ, hydrophobic effect.

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

I. INTRODUCTION

This paper reports molecular dynamics (MD) simulations exploring the conformational ensembles of the intrinsically disordered peptide Aβ as a function of variable cell size and water content, which is both of technical and chemical interest. Aβ plays a central role in Alzheimer's disease (AD) as the constituent of the hallmark senile plaques in patient brains and as the toxic oligomer-forming species whose formation is affected by mutations in genes APP and PSEN1 and PSEN2 that cause severe early onset familial AD.[1–4] Understanding the dynamics and structural transitions of this peptide under various physiologically relevant conditions is of major interest, with a particular aim of targeting the pathogenic conformations by molecular intervention such as antibodies or anti-aggregation agents.[5–10]

Molecular dynamics (MD) simulations of proteins are reported in thousands of papers each year.[11–15] Even if the chemical model and force field reflect adequately the chemical state of the protein itself, the solvent-protein system is typically studied in a periodic simulation box. It is widely assumed that image interactions in periodic MD simulations undermine results if the simulation cell is too small,[16–19] due to periodicity and artifacts in the energy evaluations. It is commonly assumed that a 10 Å distance from the edge of a box to the protein is a good practice to avoid image interactions, although this assumption is rarely tested. Diffusion coefficients[20,21] and thermal conductivities[22] may depend on simulation cell size, and water interactions are impaired in small cells.[23] In contrast, many properties have been found to be independent of cell size,[24] including to some extent protein conformation.[25] However, Hage et al. recently found that large cell sizes are required to enforce the hydrophobic effect in hemoglobin[26], but this role of simulation cell size and the water layer required to enforce the hydrophobic effect on proteins is currently debated and controversial[27,28].

Many studies have investigated the effect of Ewald summation and cut-offs of the electrostatic and the Lennard-Jones interactions[18,29–34]. Cutoff effects are, not surprisingly, particularly important for electrostatic interactions between atoms with large point charges, including charged residues of proteins[18,29,35,36]. Artificial periodicity becomes particularly problematic as the Coulomb interaction increases, i.e. for highly charged, concentrated solutes in low-dielectric solvents[18].

At the same time, one can expect the hydrophobic effect, which generally keeps the protein compact[37–39], to be impaired in small simulation cells, since a certain "critical" amount of water can be hypothesized to be required to enforce the effect. Obviously, the hydrophobic effect is a real chemical effect whereas the cell size effect is a technical effect of the MD simulation,

yet they can both work to weaken the realism of the simulation. The relative importance of the two effects is probably protein-dependent, and to estimate the critical amount of water required to keep a protein folded, intrinsically disordered proteins are ideal as they are particularly prone to conformational changes and thus likely to put an upper limit to both effects. Despite the major progress in studying intrinsically disordered proteins using MD[4041–47], we are not aware of any previous investigation of the cell-size effect on the conformational ensemble of an intrinsically disordered protein.

To understand these issues in detail, we study here a small, charged peptide that has a particularly context- and concentration-dependent ensemble: $A\beta_{40}$ is an intrinsically disordered peptide ideal for this purpose. It carries a charge of -3 at physiological pH and is extremely aggregation prone, and its secondary structure depends on the surroundings[45,48,49], plausibly enabling a conformation change from compact globular in solution to extended helical in membranes[50–53]. The properties of Aβ are very conformation-dependent and the biologically relevant conformations are very hard to identify.[50,51,54] The structural ensemble of $A\beta_{40}$ is important because its aggregation and membrane interactions are main culprits of AD,[9,10,55–57] and exposed hydrophobic parts and coil character correlate with the toxicity of Aβ variants that cause disease,[58–60] and may enhance oligomerization[10,50,61,62] or interact maliciously with cell membranes.[63–66]

MD simulations have helped to relate the structural states of Aβ to its potential biological activity.[59,60,67–74] Despite these efforts, the bioactive conformations of Aβ in human brains remain unknown, although they probably involve hydrophobic exposed parts that correlate with toxicity.[54,58,60] Aβ is known to concentrate at the synaptic cleft, the center of neuronal transmission and an extremely crowded molecular environment, with a typical width of the cleft being 20 nm.[75,76] This implies that only 60-70 aligned water molecules separate the two synaptic terminals; in this massively crowded region Aβ is likely to display its main normal and pathogenic functions.[77–80] We expect that the hydrophobic effect and its requirements in terms of water are very relevant under such conditions.

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

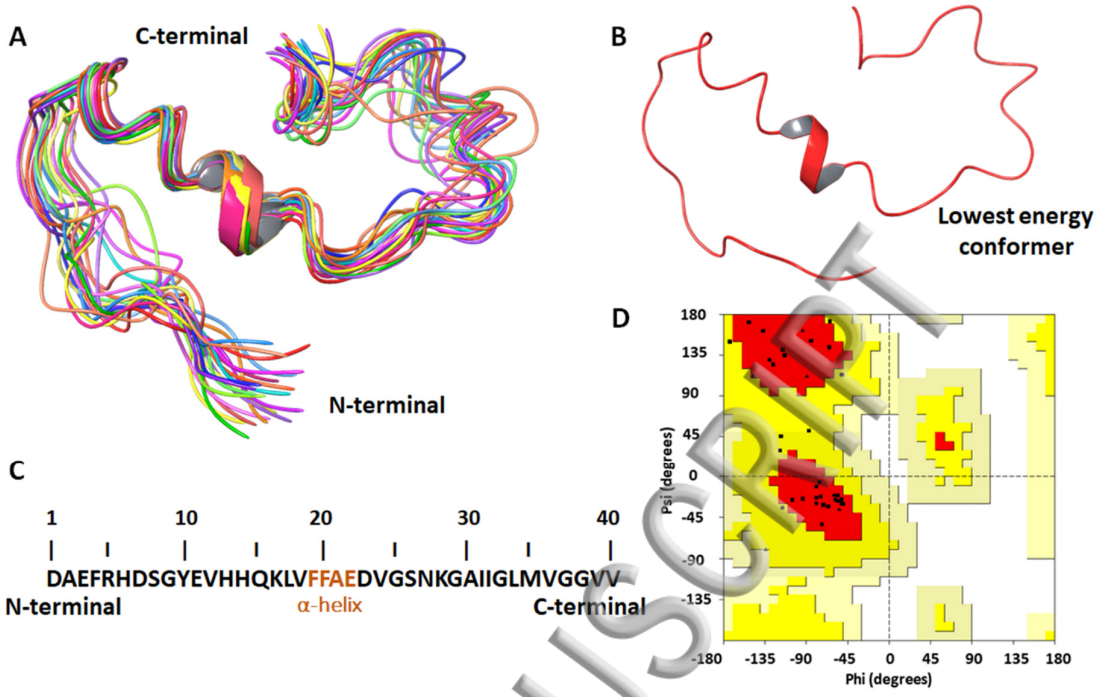


Figure 1. A) Structurally aligned twenty conformations of $A\beta_{40}$ from the PDB structure 2LFM. **B)** Lowest energy conformation (top) of $A\beta_{40}$ used in the present study. **C)** Amino acid sequence of $A\beta_{40}$ showing the residues forming α-helix. **D)** Ramachandran plot of the lowest energy conformation, with 25 residues in the core region (red) and seven residues in the additional allowed region (yellow). No residue is present in the disallowed region (white).

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

II. METHODS

A. Starting Aβ$_{40}$ model

The experimental NMR structure of Aβ$_{40}$ in aqueous solution (PDB code 2LFM; **Figure 1**)[48] was used as it represents the peptide in 100% water. Other NMR-derived structures have co-solvents or co-solutes. Starting from an extended chain is unlikely to ever reach the experimentally observed disordered state as it requires full *ab initio* folding (millisecond time scale) and such ensembles can carry bias towards structure types that have lower barriers from the extended conformation, notably coil and strand, making 2LFM a best choice of starting structure. The PDB file contains 20 conformations ranked based on lowest energies, with structural RMSD 1.7-3.7 Å (**Figure 1A**). The lowest energy conformation was used in the present study (**Figure 1B**). The peptide was prepared using Protein Preparation Wizard with default settings[81] at pH 7. Hydrogen atoms were added, hydrogen bonds were optimized and a local protein optimization was performed to remove steric clashes. The prepared protein contains -3 charge at physiological pH (please see the amino acid sequence, **Figure 1C**), with the N- and C-terminals charged as ammonium and carboxylate groups, respectively. The structure is in good agreement with available experimental NMR and CD data when using structure-balanced force fields to estimate ensemble properties[82], and has a backbone conformation without disallowed regions (**Figure 1D**).

B. System preparation

In order to understand the effect of the cell size on MD simulations, five different orthorhombic periodic cells, ranging from the commonly used (10 Å) to large (15 and 20 Å) and very small (3 and 5 Å), were studied. These cell sizes were defined by using the shortest distance from the cell edge to the peptide. The large cells test the sensitivity of the peptide properties to cell size beyond the typically used size of 10 Å, to test if this common assumption is valid for this highly conformation-sensitive peptide. In contrast, the small cells have the distance between the protein molecule and box edge below the interaction cutoff, and thus estimate the influence of interactions between the periodic images of the boxes and the removal of water layers around the peptide. The importance of these effects are likely protein dependent and thus important to know for small intrinsically disordered peptides such as Aβ.

Previously, we observed that a combination of Charmm22* force field[83] and TIP3P water model[84] produces conformational ensembles in excellent agreement with experimental NMR

and CD data, which is important because of the major effect of force field on the structure of Aβ[82]. The more polar TIP3P water model tends to favor more compact folded structures, consistent with a larger hydrophobic effect of the stronger solute-solute TIP3P interactions vs. other water models[82]. Therefore, we studied the systems using these force fields. Each system was neutralized by adding 3 Na^+ ions and additional 0.15 M NaCl was added in each. We used the viparr.py and build_constraints.py Python scripts of the Viparr tool and the System Builder tool of Desmond[81,85] to prepare the systems.

The five systems were prepared with their specific concentrations of $A\beta_{40}$ as listed in **Supplementary Table S1**. The cell volumes of the 3-, 5-, 10-, 15- and 20-Å cells are 46045, 63490, 124891, 216508 and 344340 $Å^3$, respectively. There is approximately 10 times more water molecules in the largest cell (20 Å, 10607 water molecules) than in the smallest cell (3 Å, 1097 water molecules). The concentrations range from about 0.005 M to 0.04 M. Please note that most MD simulations in the literature are in the middle range of this interval as they commonly use box sizes of 10 Å[82], whereas experimental assay concentrations are much more dilute, in the micromolar range[77,86]. It is thus important to know if this approximation done in very many studies has a negative impact on realism vs. experimental assays.

C. Energy minimizations and MD simulations

Each system was energy-minimized using a combination of steepest descent and Broyden−Fletcher−Goldfarb−Shanno methods. MD simulation was then performed in NPT ensemble at 300 K and 1.0013 bar for 100 ns using a multistep protocol of Desmond[81,85]. The integration time-step of 2 femtoseconds was used. The energies were noted with an interval of 1.2 picoseconds (ps) and trajectories were recorded at each 100 ps interval that generated 1001 snapshots in each run. The temperature and pressure were kept stable using the Nose-Hoover chain thermostat[87] and the Martyna-Tobias-Klein barostat[88]. The long-range electrostatic interactions were calculated using Ewald mesh summation technique with a cut-off of 9.0 Å. Each simulation performed in this study was initiated using a different, randomly seeded start velocity. The initial velocities for each atom were drawn randomly from a Maxwell-Boltzmann distribution. In total, 10 μs of these concentration-dependent all-atom simulations were performed with the Charmm22* force field, divided into 100 seeded 100-ns simulations. Additionally, 16 μs simulations were computed with two other force fields as described below.

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

Additional ten randomly seeded simulations with 5 Å cell size were performed for 100 ns at 320 K with all other parameters similar to the other simulations, in order to analyze the effect of varying temperature on both sides of the physiological regime (most assays are performed at room temperature, 300 K, but the peptide is active at body temperature, 310 K). We hypothesized that thermal disorder could affect the peptide dynamics in particular in the low water concentration limit where the water-water interactions are weaker.

To understand the impact of sampling using many short (100 ns) vs. a few longer (1000 ns) MD simulations, we also performed three randomly seeded simulations of the system with 10 Å cell size, each lasting for 1000 ns at 300 K under same conditions as above except for the trajectory recording time. These trajectories produced a substantially longer averaging history relevant to the identification of slow processes with considerable diffusion barriers. Specifically, we wanted to test if there were any differences in the buildup of strand character at long simulation times to a level that could not be reached on the 100-ns time scale. Each 100 ns trajectory was analyzed for the equilibrated last 50 ns time, and the 1000 ns trajectories were analyzed for the last 950 ns and 500 ns time, to avoid artifacts of the first equilibration phase where structures diverge from the minimized initial start structure. To test for long diffusion-limited processes, we averaged the long simulations both over the last 950 ns and over the last 500 ns and analyzed the differences compared to the short simulations.

Finally, to test the sensitivity of our results to the choice of force field, we performed twenty simulations for each of the 3-, 5-, 10-, and 15-Å cells using two additional force fields, OPLS3e[89,90] and Amber99sb-ILDN[91]. Each simulation was performed for 100 ns using the same protocol as described above (for a total of 8 μs with each force field). Previously, Charmm22* and Amber99sb-ILDN were shown to be accurate and structure-balanced for studying Aβ ensembles[82], and we thus expect the results provided here to be maximally realistic in terms of force field choice.

The secondary structures, RMSD and RMSF plots were generated using the Simulation Interaction Diagram tool of the Desmond, and the hydrogen bonds and radius of gyration were studied using Simulation Event Analysis tool. For analyzing the hydrophobic and the hydrophilic SASA, the Python script trj_sasa.py was used with a probe radius of 1.4 Å, and the representative structures of the trajectories were identified by affinity propagation using the trj_cluster.py script of Schrodinger with default parameters[81].

III. RESULTS AND DISCUSSION

A. Sampling quality of MD simulations

We performed MD simulations of $A\beta_{40}$ using simulation cells of five different sizes (3 Å, 5 Å, 10 Å, 15 Å, and 20 Å), with both two smaller and two larger cells than the typical size used in default studies (~10Å). For each cell size, we performed 20 randomly seeded simulations for 100 nanoseconds, giving 2000 ns for each cell size. We also performed three longer simulations of 1000 ns each to test the impact of averaging over long vs. short MD simulations. Furthermore, we studied the peptide for additional 10 x 100 ns at 320 K to understand the sensitivity of our findings to reasonable variations in temperature. We analyzed the secondary structure, radius of gyration, hydrophilic and hydrophobic solvent accessible surface area (SASA), inter- and intra-molecular hydrogen bonding.

To estimate whether MD simulations adequately sample a property of interest, one should compare many simulations with different randomly seeded start velocities, run these simulations for a considerable time, and then consider them together as independent observations[92]. Only the later horizontal parts of the trajectories are included in the statistics, as the initial trajectories reflect divergence from the starting structure, which is correlated to snapshots of the early trajectory. The RMSD and RMSF plots of all trajectories are shown in **Supplementary Tables S2-S13**. In contrast to many systems and experiments where precision is usually quite high and triplicates are enough to reach decent results, Aβ is fundamentally disordered and has a large variance in its conformational ensemble, which produces stochastic (chaotic) results even in experimental assays.[77]

Similarly, very large standard deviations can be expected in MD for observables with large amplitude of motion even if they are well-sampled; this is true for conformational properties of Aβ as these are naturally fluctuating on fast timescales[82]. Some properties with slow time scales (> 1 microsecond) may be completely outside the simulation time scale, most notable changes in secondary structure of stable folded proteins, which have large diffusion barriers to change structure[19,93]. Not surprisingly, in disordered proteins such as Aβ, the diffusion barriers to changes in secondary structure are much smaller, i.e. in nanoseconds timescale. We show below that longer (1000 ns) and shorter simulations (100 ns) produce similar prevalence of secondary structure transitions, and thus the characteristic time for these events is < 100 ns for this disordered peptide, much shorter than for compact folded proteins with large barriers separating the secondary structure types. Accordingly, as analyzed below, all the properties of interest in this study are well sampled on the applied simulation time scales.

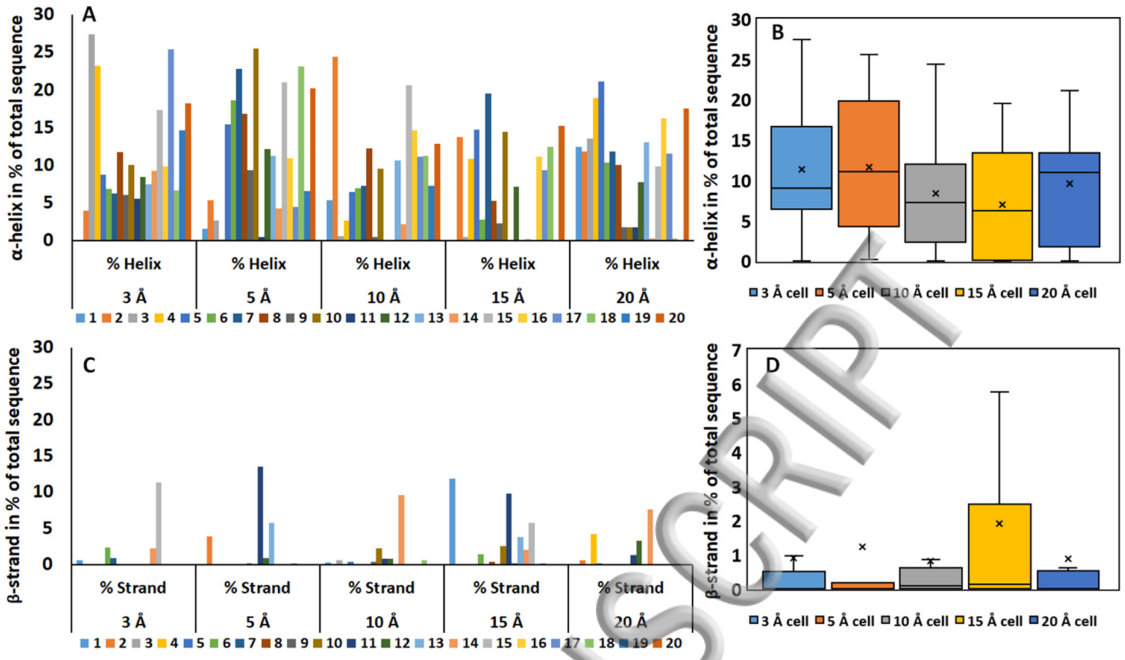


Figure 2. A) Percentage of residues with α-helix structure for 100 different 100-ns simulations of Aβ$_{40}$, with 20 simulations at five different cell sizes (3, 5, 10, 15, and 20 Å) (Charmm22*; averaged over the last 50 ns of each simulation). **B)** Box-and-whisker plot of the helix content shown in A. **C)** Plot (as in A) of the percentage of β-strand. **D)** Box-and-whisker plot of strand content varying from 0-6%.

B. Disordered Aβ with ~0-27% (average 8-12%) helix and common α-β transitions

The analysis of the secondary structures in all the simulations at 300 K (with different cell sizes: 3, 5, 10, 15 and 20 Å) shows that α-helix was persistent except for a few simulations where β-strand was prominent (**Figure 2A-2D** and **Supplementary Tables S14-S20**). In all cases studied, the peptide remains predominantly disordered, with most of the residues being coil, turn or bend, in accordance with the experimental data.[48,82] Complete loss (< 2%) of secondary structure, with neither helix or strand present, occurred regularly (13 times out of 100 simulations) and was independent of cell size.

In most simulations, some helix was present, averaging to ~8-12% for the five cell sizes, with standard deviations of 6-8% helix character (**Supplementary Table S20**). The average strand character was 1-2%. Typically, β-strand formation to 5-15% was observed in 1-2 seeds of each of the five sets of 20 seeded 100-ns simulations, making α-to-β transition a relatively common feature (5-10%) of the peptide at all water concentrations. None of these tendencies

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

were significantly dependent on the cell size or water content, and the similar tendency of α-to-β transition in all five groups suggest that the initial process of this transition is well-sampled. In the 15-Å cell, two larger (10-12%) α-to-β transitions and two smaller (4-6%) produced a larger standard deviation (SD) in the strand content for this peptide but these variations are still less than 6% and should be seen in the context of the very small average strand content. In conclusion, the secondary structure and initial events of the α-to-β transitions are well-sampled and independent of simulation cell size. Even at very small cell sizes (3 and 5 Å) where image interactions are clearly involved, the secondary structure content was similar to that of the larger cell sizes (10, 15 and 20 Å).

C. Hydrogen bonds and backbone conformations of Aβ at variable water concentrations

In order to analyze the hydrophobic packing of $A\beta_{40}$, intra-peptide and peptide-water hydrogen bond networks were analyzed (**Figure 3A,** and **Supplementary Table S20**). With an increase in the hydrophobic packing, the peptide-water hydrogen bond networks will decrease and intra-peptide hydrogen bonding will increase. For the 3-Å cell, the number of intra-peptide and peptide-H_2O hydrogen bonds lies in the intervals 16.2−25.3 and 100.6−118.7, respectively, with averages of 20.0 (SD = 2.7) and 111.6 (SD = 5.0), respectively. In the 5-Å cell, the number of intra-peptide and peptide-H_2O hydrogen bonds is 15.9−24.6 and 102.8−121.2 for the simulations, and their averages over all simulations are 20.7 (SD = 2.9) and 111.1 (SD = 4.9), respectively. Similar patterns prevail for the 10-Å, 15-Å, and 20-Å cells. Correspondingly, the peptide-water hydrogen bonds lie between 105.1−122.0, 101.3−125.9 and 104.7−123.0, with averages of 114.8 (SD = 4.5), 115.3 (SD = 6.9) and 113.2 (SD = 5.5) for the 10-Å, 15-Å, and 20-Å cells, respectively. We thus conclude that the number of hydrogen bonds is unaffected by the cell size, i.e. the number is similar for both very large (with high water content) and very small (with less water content) cells.

To understand the impact of the reduction in water potential on the backbone conformations of the peptide, the Ramachandran plots of the representative structures of each simulation from cluster analysis were also analyzed (**Figure 3B** and **Supplementary Table S20**; excluding the six glycine residues and two terminal residues, as they can occur in any region of the plot). For comparison to experimental data, the 20 NMR conformers from the 2LFM entry[48] were also analyzed. These conformers have 71.9−90.6% of the residues (23-29 residues) in the core region and 9.4−28.1% (3−9 residues) of the 32 analyzed residues in the additional allowed region.

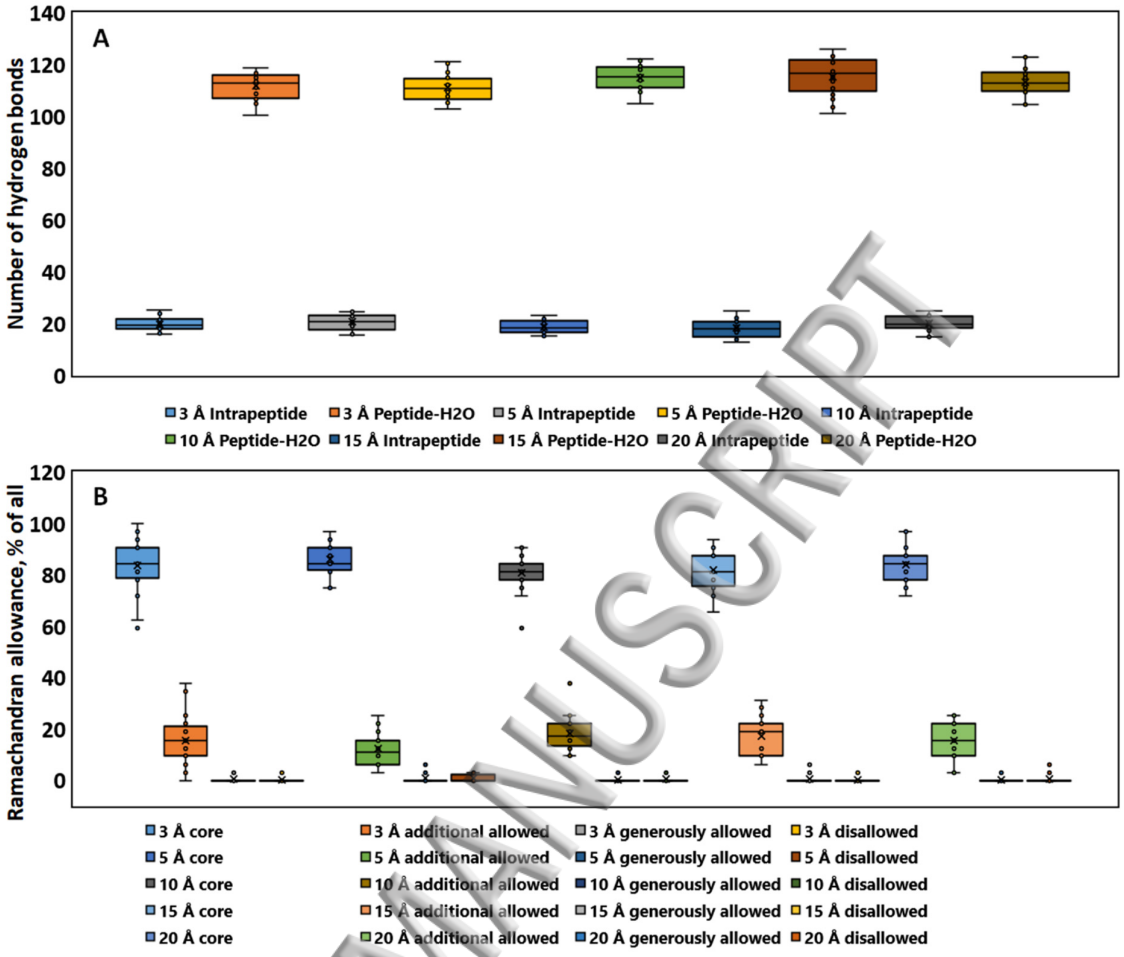


Figure 3. A) Box-and-whisker plot of the number of peptide-water hydrogen bonds and peptide-peptide hydrogen bonds for the last 50 ns of 100 different 100-ns Charmm22* simulations. The top values represent peptide-water hydrogen bonds and the bottom values represent peptide-peptide hydrogen bonds. **B)** Box-and-whisker plot of backbone conformations assigned to core, allowed, generously allowed and disallowed regions of the Ramachandran plot, averaged over the last 50 ns of all 100 100-ns simulations.

In most of the representative structures for all cell sizes, no residues had disallowed backbone conformations, except for few simulated structures, where one or two residues (two residues only in two cases) were present in the disallowed region. Notably, in the smallest cell (3 Å), the variations in the core and additional allowed regions were consistently larger, a feature that relates to the high variation in the hydrophobic SASA. This may be due to the constraint in the conformational relaxation and the periodic image interactions in very small cell size with only one or two water layers. However, it also shows the reduced ability of a few

water layers to keep the peptide compact. The 15-Å cell was an exception, caused by the high frequency of α- to -β transitions. Nevertheless, the Ramachandran values of all cells are in excellent agreement with the 20 NMR conformers of 2LFM. The average core regions for 3, 5, 10, 15 and 20 Å cells were 83.5% (SD = 10.1), 86.2% (SD = 5.8), 80.9% (SD = 7.0), 81.9% (SD = 7.4) and 84% (SD = 6.5), respectively.

D. Solvent accessible surface area

In order to understand the early conformational changes of $A\beta_{40}$ occurring at low water potential, the hydrophobic and hydrophilic solvent accessible surface areas (SASA) of the peptide were analyzed for all 20 simulations of each of the five cell sizes (**Figure 4A** and **Supplementary Table S20**). It is interesting to note that Aβ has very similar total hydrophobic and hydrophilic exposure, both in the range 1500−2000 $Å^2$, indicating its very high amphiphilicity, a feature that rationalizes its well-established interaction with membranes[49,52,94].

We observe interesting trends in the ensemble-averaged surface behaviors across the 100 simulations (**Figure 4A**). First, the hydrophobic SASA is significantly more variable in the two smallest cells where only a few layers of water are generally available. The average hydrophobic SASA in the small cells is also comparatively higher, 1760 $Å^2$ (SD = 271 $Å^2$) and 1718 $Å^2$ (SD = 272 $Å^2$) for the 3-Å and 5-Å cells, compared to the values 1598 $Å^2$ (SD = 186 $Å^2$), 1582 $Å^2$ (SD = 194 $Å^2$) and 1628 $Å^2$ (SD = 121 $Å^2$) of the 10-Å, 15-Å, and 20-Å cells (**Supplementary Table S20**). In all the larger cells, the hydrophilic surface area consistently separates from the hydrophobic surface area, whereas in the 3-Å and 5-Å cells, the hydrophobic exposure variability completely overlaps the range of hydrophilic exposure and thus routinely turn the balance in favor of hydrophobic exposure.

The data for the large cells show that 10 Å is sufficient for all studied properties of Aβ, largely validating the use of modest-sized simulation cells for MD simulations of proteins that are less structure-sensitive than Aβ. In contrast, the small cell sizes are not physically relevant as they contain periodic image interactions leading to conformational restriction effects. However, the fact that we observe similar values for almost all properties for small and large cell sizes even for a highly charged, disordered peptide such as Aβ strongly indicates that artifacts of small cells are small and potentially overestimated. On the other hand, the very central ability of the peptide to stay compact (the hydrophobic effect) interestingly plays out in

the range of typical simulation box sizes, with 10 Å being required to consistently enforce the compact state. This suggests that the hydrophobic effect is mainly enforced by water-water interactions (rather than peptide-peptide or peptide-water interactions).

Hage et al. found that large cell sizes are required to describe hemoglobin[26]. This result has been debated by Gapsys and de Groot,[27,28] and thus the role of simulation cell size and the water amount required to enforce the hydrophobic effect is currently controversial. We expect that the hydrophobic effect requires more water layers for a large solvent-exposed surface such that the critical amount of water needed to enforce the hydrophobic effect is protein-size dependent. Whereas large cells may be needed for hemoglobin[26], only 10-Å cells, or 2-3 layers of water molecules, are required to enforce the hydrophobic effect of Aβ. The hydrophobic exposed parts of Aβ probably represent the bioactive conformations and correlate to toxicity[51,54,58]. The synaptic cleft where Aβ concentrates[79] only has 60-70 water molecules between the synaptic terminals of two neurons,[75,76] and we thus predict that impaired hydrophobic effects will be an important feature of peptide dynamics and lead to hydrophobic exposure *specifically* in this region where Aβ is active, which should be very relevant to studies of Aβ toxicity and drug development efforts.

We conclude that Aβ has a smaller tendency to exhibit hydrophobic exposure at high water potential, and that the hydrophilic exposure outweighs the hydrophobic exposure, but that the amphiphilic balance is destroyed at low water potential. The differences are statistically significant, as shown from a two-tailed Student t-test for hypothesized same mean. Whereas the 3-Å and 5-Å cells have insignificantly different hydrophobic SASA (**Supplementary Table S21**), the mean hydrophobic exposure is significantly different at the 95% confidence level for the small vs. all the larger 10-, 15-, and 20-Å cells when these tests are performed independently (**Tables S22-S24**). In contrast, the hydrophilic surface is not significantly related to cell size (**Tables S25-S28**), and the average values are very similar with 1840 Å^2 (SD = 142 Å^2), 1846 Å^2 (SD = 117 Å^2), 1791 Å^2 (SD = 125 Å^2), 1816 Å^2 (SD = 163 Å^2), and 1779 Å^2 (SD = 119 Å^2) for the 3-Å, 5-Å, 10-Å, 15-Å, and 20-Å cells, respectively.

We hypothesized that this significant change in the balance between hydrophilic and hydrophobic exposure in small cells may be accompanied by a change in the size of the peptide. To test this, the radius of gyration (R_g) of $A\beta_{40}$ in all the simulations of different cell sizes was computed. R_g is very similar ranging across all simulations from 9.4 to 13.8 Å with average R_g ~11 Å and SD 0.6-1.0 Å (**Figure 4B** and **Supplementary Table S20**). The change in cell size does not have a statistically significant effect on R_g of the peptide. However, the within-

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

ensemble averages of R_g and hydrophobic surface exposure correlate significantly, as the exposure is related to a partial increase in the size of the peptide, and for all five cell sizes the relation has R^2 values between 0.35-0.50 (**Figure 4C** and **Supplementary Figure S1**).

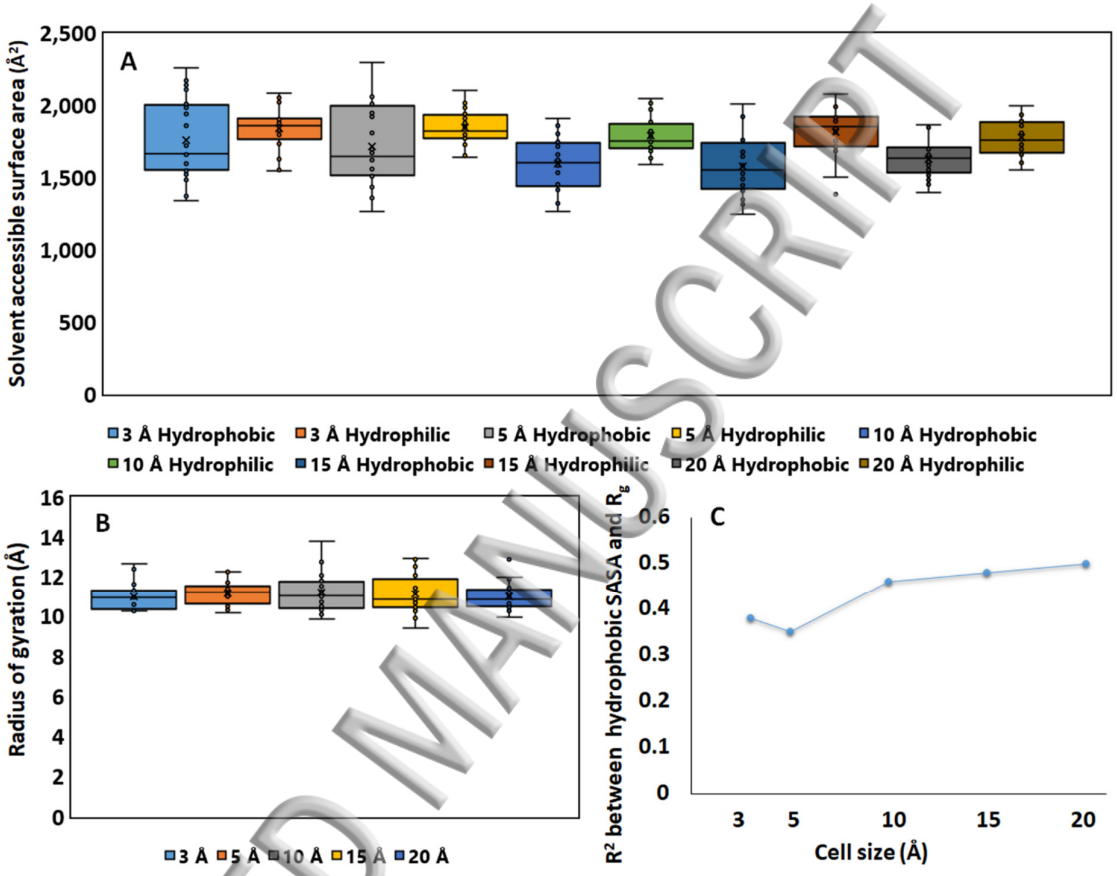


Figure 4. A) Solvent-accessible surface area of Aβ, divided into hydrophobic and hydrophilic parts, averaged over 20 MD simulations at five different simulation cell sizes (3, 5, 10, 15, and 20 Å). **B)** Box-and-whisker plot of the radius of gyration (R_g) of the simulations at five cell sizes. **C).** Calculated R^2 from linear regression of the relation between R_g and hydrophobic solvent-accessible surface area for 20 simulations at each cell size; in all cell sizes, the two properties correlate significantly (95% confidence interval) with R^2 values of 0.35-0.50.

E. Terminal distance analysis and sampled Aβ structures

We also analyzed the distance between the Cα-atoms of the N- and C-terminal residues (**Figure 5A** and **Supplementary Table S29**) of the representative structures from cluster analysis at each cell size. The average distances are 24.6 Å (SD = 11.8 Å), 24.9 Å (SD = 14.0 Å), 20.9 Å

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

(SD = 7.7 Å), 20.4 Å (SD = 4.9 Å) and 18.9 Å (SD = 7.6 Å) for the 3-, 5-, 10-, 15-, and 20-Å cells, respectively. The average distances and their standard deviations are consistently larger for the small cells (3 and 5 Å) than for the large cells (10, 15 and 20 Å). These terminal distances relate to the hydrophobic effect, since the distance correlates with the hydrophobic SASA. **Figure 5B-5G** shows representative structures for the different cell sizes indicating this relationship.

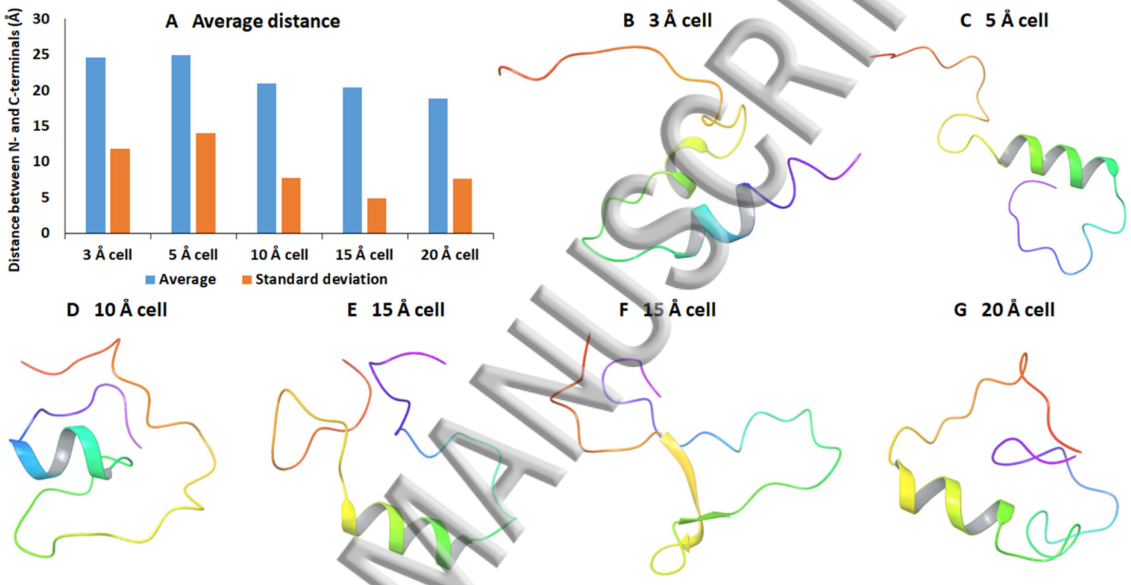


Figure 5. A) Average and standard deviation of the distance between the Cα-atoms of the N- and C-terminal residues of $A\beta_{40}$ for each cell size. **B-G)** Representative structures from cluster analysis of the twenty Charmm22* simulations of each cell size, except **F)** showing a snapshot of the 15-Å cell size with high β-strand occurrence as example of an α-to-β structural transition.

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

F. Regional variations in secondary structure

The experimental NMR structure 2LFM contains 20 lowest energy conformers of $A\beta_{40}$.[48] All these conformers possess 10% α-helix (Phe-19 to Glu-22) (**Figure 1**) and no strand. In addition, they also comprise 15% 3_{10} helix (His-13 to Val-18), which are not very stable in water.[95] To avoid biases from the assignment, we focused our analysis on the most stable elements, i.e. all α-helices and β-strands. The percentage of α-helix in the experimental structure 2LFM is comparatively similar to the average α-helix for each cell size (**Figure 6A**), suggesting that our structural ensemble is in good accordance with the NMR data in water. Other NMR-derived structures of the Aβ monomer in PDB display more helix but they reflect the presence of co-solvents and micelles that increase the helix character[77]. Other data, notably coupling constants and chemical shifts for the peptide in pure water, also support that the peptide is mainly disordered with a small amount of helix and little strand.[82,96] In contrast, the strand content rises quickly in the dimers and larger oligomers and is very large in the fibrils that form the senile plaques of AD[54,97].

Apart from the α-helix in the experimental structure (Phe-19 to Glu-22), we also identified three more regions forming helices during the simulations (**Figure 6B, left** and **Supplementary Table S30**). We divided the residues into four helix-forming regions: Region 1 covers residues 1-10, region 2 covers residues 11-20, region 3 is from 21-26 and region 4 covers residues ~27-38. Regions 3 and 4 were also prominent at all cell sizes, particularly region 4. Furthermore, as the cell size increased up to 10 Å, region 2 (experimental helix) occurred 7 times and region 4 occurred 11 times. Region 1 occurred very rarely at all cell sizes.

We also divided the residues into three strand regions: 1 (residues 1-10), 2 (residues 11-20), and 3 (residues 21-40) (**Figure 6B, right** and **Supplementary Table S30)**. All these regions occurred very rarely. In the 15-Å cell, β-strand occurred most frequently (11 times). Regions 1, 2 and 3 occurred three, four, and four times, respectively (**Figure 6B**). Thus, apart from the experimentally determined helix region[48], helix regions 3 and 4 have very high tendency to occur as α-helices; these are, interestingly, the regions that are experimentally observed in structures such as 1BA4 and other NMR-derived structures in the PDB where the water potential has been reduced[98]. Overall, we conclude that, as summarized by **Figure 6**, the helix and strand contents of all simulations, even when divided into segments of the peptide sequence, are in good agreement with the experimental 2LFM conformations and are not significantly affected by the cell size. This means that the secondary structure contents are similar even in the smaller (3 and 5 Å) and the larger (15 and 20 Å) cell sizes.

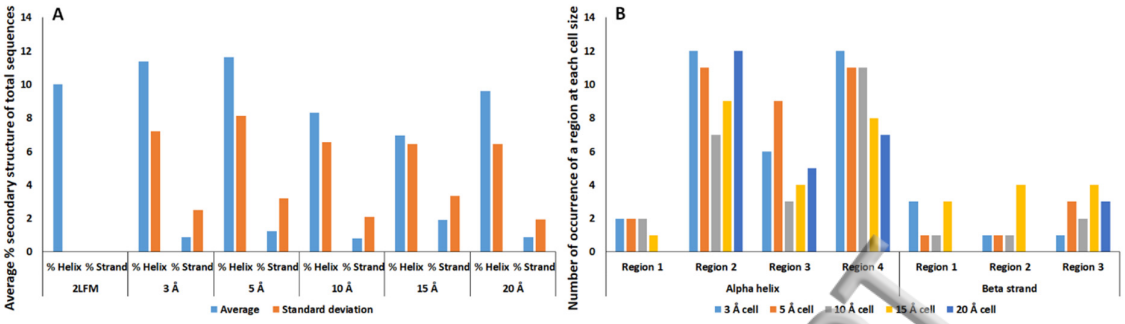


Figure 6. A) Average percent secondary structure found in each cell size compared with the average of 20 conformers of experimental NMR structure 2LFM. **B)** Analysis of the regions forming α-helices and β-strands. Number of times a region forms an α-helix or a β-strand in the 20 simulations of each cell size. The secondary structural regions that occurred ~ >20% of the analysis time (last 50 ns) of each run were analyzed.

G. Results are not sensitive to variations around physiological temperature

The hydrophobic effect is related to temperature and manifests as an entropy-driven exposure of proteins and peptides as the water network is weakened. We thus hypothesized that the temperature could affect the observed exposure. To ensure that our findings are not sensitive to reasonable temperature changes near physiological temperature, we performed 10 x 100 ns of additional simulations for the second-smallest (5-Å) cell at 320 K, which is 10 K above physiological temperature, with the other simulations performed at 300 K, 10 K below body temperature, and typical for experimental assays of Aβ.

We compared this 320-K simulation with the 300-K simulation of the same 5-Å cell system (**Supplementary Table S20**). At 320 K, large variability is seen in all properties to the same extent as at 300 K. The increasing temperature in principle causes thermal disorder in the water-water interactions which could affect our conclusions, but the variability is similar at both temperatures and robustly captured by the averaging of the ensemble properties. R_g was insignificantly different at 300 and 320 K: At 320 K, R_g lies between 10.3-12.2 Å, with average 11.3 Å and SD 0.7 Å. The intra-peptide (average = 20.6, SD = 3.4) and peptide-water hydrogen bonds (average = 108.2, SD = 6.3) were insignificantly affected by the higher temperature. Most importantly, the increasing temperature had no effect on the hydrophobic SASA in the range around physiological temperature (**Supplementary Table S20**), and thus, our observations of a hydrophobic exposure at low water potential is robust against variations in temperature in the range relevant to assays and physiological conditions.

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

H. Averaging data for long versus short simulations

It might be argued that some of the slowest modes in the ensemble, notably the α-to-β transitions, occur on longer timescales than studied here, and that they require more time to travel up the diffusion barrier; this argument rests on the assumption that disordered peptides have slow diffusion barriers for these transitions as commonly seen for compact folded proteins[19]. However, we expect disordered proteins to have much smaller barriers to these transitions and thus a shorter time scale than compact folded proteins. In order to investigate this, we compared averaging for long versus short MD simulations by performing additionally three 1000 ns simulations and comparing these to the twenty 100 ns simulations using the 10-Å cells. For the long simulations, the data were additionally averaged both over the last 950 ns and the last 500 ns to detect any timescale-dependence of the properties and their variations (**Figure 7**).

One long simulation produced more strand than the average of the 20 100-ns simulations. In this simulation, 5.2% and 9.9% strand occurred in the last 950 ns and 500 ns of the 1000 ns, indicating a buildup and persistence of strand character (**Figure 7A** and **Supplementary Figures S2** and **S3**). The percent α-helix in this long simulation was comparable to the average of the short simulations. 13.9% and 10% α-helix were found during the last 950 ns and 500 ns, compared to 8.3% average α-helix (SD = 6.5) in the short simulations. In contrast, the two other long simulations exhibited secondary structures very similar to the average of 20 short simulations. The percentage of α-helix and β-strand falls in the intervals 5.4−6.4% and 0.5−2.7% during the last 950 ns and 5.0−7.2% and 0.1−1.2% during the last 500 ns, respectively. The amount of strand buildup during the first long simulation is not higher than seen in the short simulations, and the variance in the properties is similar.

From this, we conclude that α-to-β transitions are well-sampled to the same extent by the long and short simulations but still require a total simulation time of ~1000 ns to ensure its observation. These transitions occur 5-10% of the simulation time, build up to ~10% regardless of time scale, and thus have much smaller diffusion barriers than larger folded proteins. This shows that the timescale and thus length of MD simulations required depends not only on the type of transition but also on its free energy barrier, which can be very small for disordered proteins and peptides. In larger compact folded proteins, the secondary structure elements are longer and tend to pack in tertiary structure, which stabilizes greatly the secondary structure and produces large barriers for structure transitions, in direct contrast to our findings for Aβ.

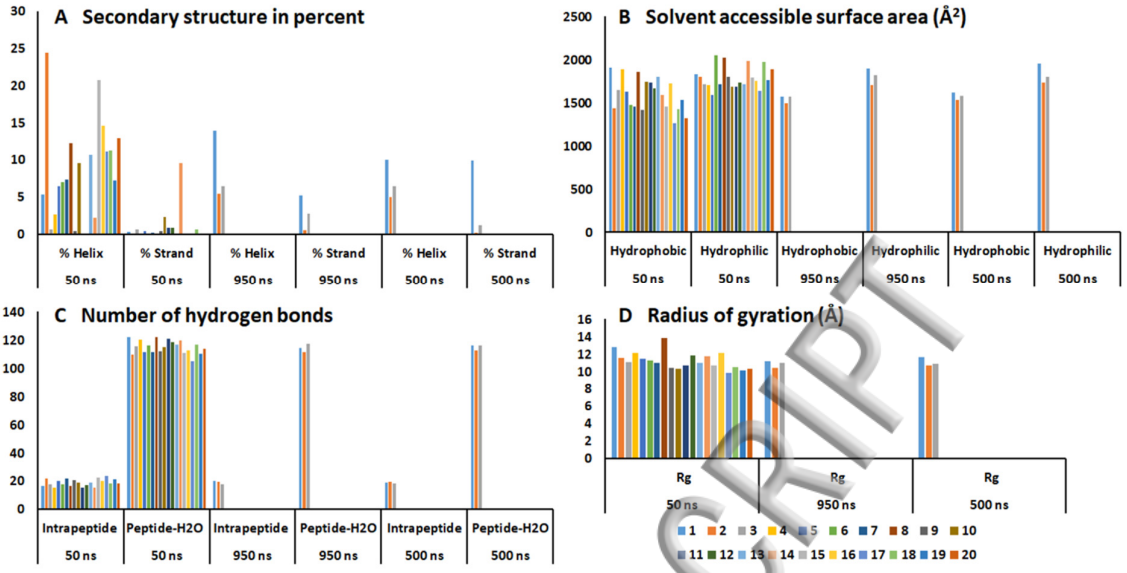


Figure 7. MD simulation results analyzed over 50 ns of the 100 ns runs (20 simulations) and over 950 ns and 500 ns of the 1000 ns runs (3 simulations). **A)** Secondary structure in percent. **B)** Hydrophobic and hydrophilic solvent accessible surface area (Å^2). **C)** Number of intra-peptide and peptide-water hydrogen bonds. **D)** Radius of gyration (Å).

The hydrophobic exposure is a much faster event but its prevalence could be biased by poorly sampled larger conformational changes. To support that our main conclusions on the hydrophobic exposure of Aβ are not dependent on the time scale of simulation, the hydrophobic and hydrophilic SASA averaged over the last 950 and 500 ns showed similar values to the average of the 20 100 ns simulations (analyzed for the last 50 ns, **Figure 7B**). For the 950 ns and 500 ns averaging, hydrophobic SASA lies between 1490-1571 Å^2 and 1531-1622 Å^2, and the hydrophilic SASA lies between 1705-1899 Å^2 and 1736-1956 Å^2, respectively.

We conclude that for intrinsically disordered peptides like Aβ (not for compact proteins in general), averaging over longer or many shorter MD simulations makes no significant difference because of the small transition barriers. We stress that these conclusions necessarily rely on a fraction of the phase space of Aβ. REMD simulations can produce more β-strand[99], as the high-temperature configurations favor transitions to β-strand in the overall sampling. These simulations produce similar high (25% and 22%) strand content for OPLS-2001 and Amber99sb-ILDN with strand in the C-terminal[99], despite their distinct secondary structure preferences in benchmarks[100,101], which is evident in conventional MD of Aβ.[82] C-terminal β-strand independent of force field use was also seen in another REMD study of Aβ$_{42}$.[102] It is possible than very slow α-β transitions exist, and that these strands are so stable that they prevail

with all force fields. Yet all available Aβ monomer NMR structure models 1BA4[98], 1IYT[103], 2LFM[48], 1Z0Q[96], and 1AML[104] indicate some helix character and very little strand in Aβ, and no strand after the turn,[105,106] as also supported by the absence of a signal at early times in thioflavin T fluorescence studies and other CD and NMR data.[48,77,82,96,107] If the peptide is oxidized, not properly monomerized, or left for hours in assays, it forms strand[96,107], making the experimental data used for comparison very important.[77]

I. Independence of results on choice of force field

Finally, we tested the sensitivity of our results to the choice of force field by performing 20 100-ns simulations for four cell sizes (3, 5, 10 and 15 Å cells) using both OPLS3e and Amber99sb-ILDN force fields (**Figure 8** and **Supplementary Tables S31** and **S32**). We did not perform simulations using the larger 20 Å cell, because we observed similar properties already for the 10- and 15-Å cells.

The secondary structure content obtained by the three force fields differed somewhat (**Figure 8A**), although not as much as for some other force fields[82]: OPLS3e formed the highest percentage of α-helix in all four cell sizes, whereas Amber99sb-ILDN produced the least helix. The difference in helix content of Charmm22* and OPLS3e was significant (two-tailed t-test) for all cells except the 5-Å cell (**Supplementary Tables S33-S36**). The helix difference of Charmm22* and Amber99sb-ILDN was also significant for all cell sizes except the 15-Å cell (**Supplementary Tables S37-S40**), and the OPLS3e produced significantly more helix than Amber99sb-ILDN for all four cells (**Supplementary Tables S41-S44**). As for Charmm22*, the β-strand content was small and insignificantly different also with these force fields (**Figure 8A**), and the number of intra-peptide and peptide-water hydrogen bonds (**Figure 8B**) and Ramachandran values also showed no significant dependency on cell size (**Figure 8C**).

The solvent accessible surface area behaved similarly as when using Charmm22*: The hydrophobic SASA became more variable in the small 3-Å and 5-Å cells in particular when compared to the 15-Å cells (**Figure 8D**). The average hydrophobic SASA was again comparatively higher for the 3-Å and 5-Å cells using OPLS3e (**Supplementary Table S31**) or Amber99sb-ILDN (**Supplementary Table S32**). The 3-Å and 5-Å cells displayed similar hydrophobic SASA using both OPLS3e and Amber99sb-ILDN, whereas the difference between small and large cells was statistically significant (95% confidence; one-tailed t-test; **Tables S45-S50**). The hydrophilic SASA showed a pattern corresponding to the hydrophobic SASA (**Figure 8D**): It was insignificantly different in 3-Å and 5-Å cells but significantly larger

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

than in the 10-Å and 15-Å cells (**Tables S51-S56**) using both OPLS3e (**Table S31**) and Amber99sb-ILDN (**Table S32**). Also consistent with Charmm22*, the radius of gyration was not affected by cell size using the two alternative force fields (**Figure 8E**), and again, the R_g and hydrophobic SASA weakly correlated, with R^2 values between 0.1–0.4 using OPLS3e, and 0.1–0.6 using Amber99sb-ILDN (**Figure 8E** and **Tables S31** and **S32**).

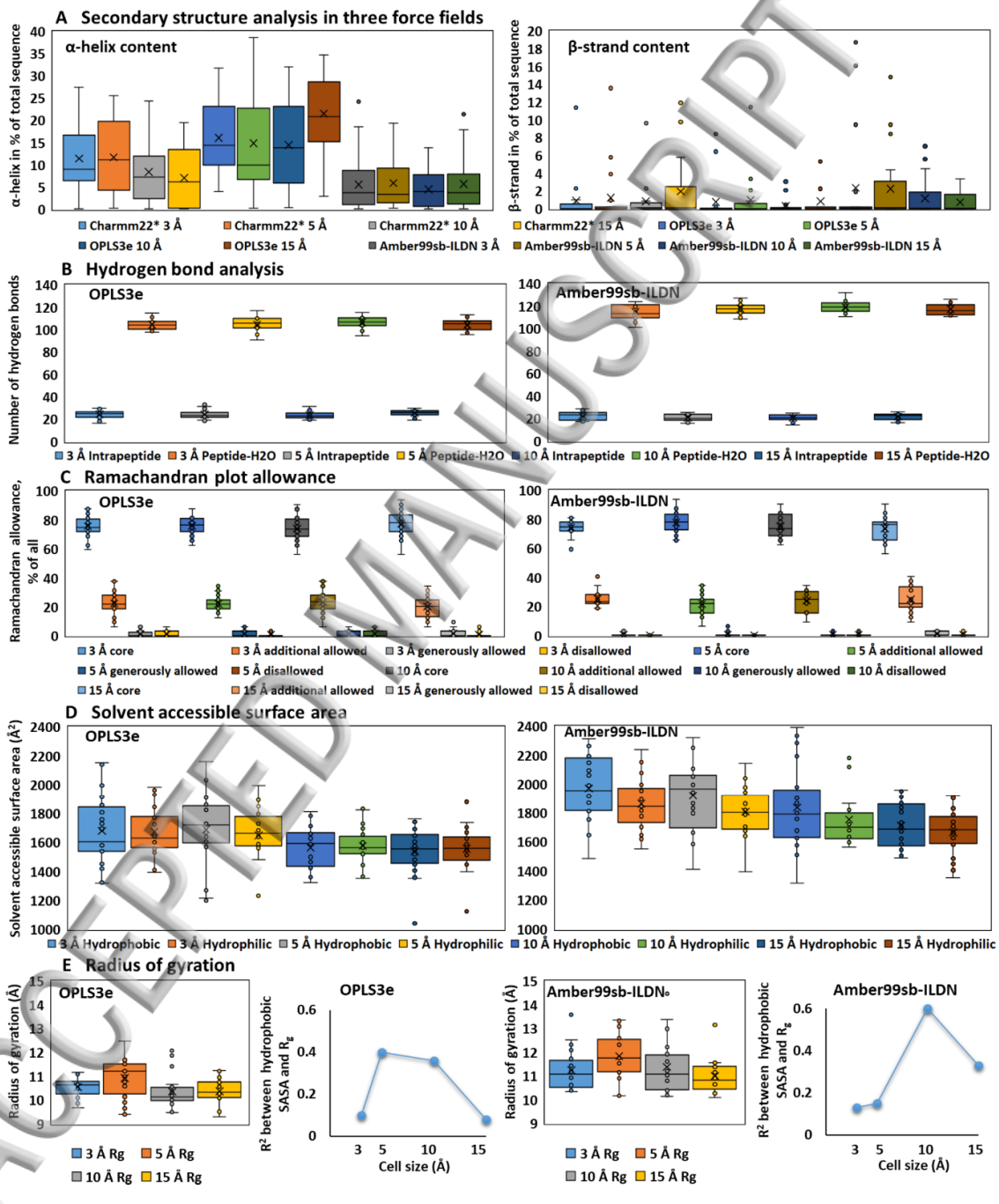


Figure 8. Analysis of the MD simulations performed using OPLS3e and Amber99sb-ILDN force fields. **A)** Comparison of secondary structure using Charmm22*, OPLS3e and

Amber99sb-ILDN. **B)** Hydrogen bond analysis. **C)** Ramachandran analysis. **D)** Solvent accessible surface area. **E)** Radius of gyration and its correlation with hydrophobic SASA.

In conclusion, despite differences in secondary structure propensity of the force fields, the simulations with Amber99sb-ILDN and OPLS3e produced very similar cell-size dependencies of the MD-simulated properties as Charmm22*. Accordingly, our distinction between cell-size dependent and independent properties is not much affected by the choice of force field, making the requirement of only 2-3 layers of water molecules to obtain size-consistent results for Aβ robust to the choice of force field.

IV. CONCLUSIONS

Almost all commonly performed MD simulations of proteins are extremely concentrated compared to experimental assay conditions, and in addition to sampling and force field quality, the chemical composition may thus be a major issue in MD simulations using typical cell sizes of 10 Å. To understand this we studied 15-Å and 20-Å cells in comparison to 10-Å cells for the very disordered and highly charged peptide Aβ, which we expect to be particularly sensitive to cell-size effects and thus represent a worst-case example of such effects.

We show here using the structure-balanced force field Charmm22*[83] that even for Aβ, the cell size from 10 to 20 Å is unimportant for many properties typically studied by MD simulations. This includes the radius of gyration, hydrophilic SASA, intra-peptide and peptide-water hydrogen bonding, and secondary structure. Since Aβ is very disordered, we expect the same size-insensitivity to manifest for most other peptides and proteins. By comparing long and short simulation and different averaging, and using statistical tests for means, we show that our simulations sample all the properties well and thus recover the intrinsic variability of the peptide, which is very large, rather than variability due to statistical uncertainty. Our simulated structures are in excellent agreement with the NMR-derived conformations in pure water reflected by the structure 2LFM.[48]

However, we also see that the hydrophobic exposure was significantly larger in small cells (3 and 5 Å) where image interactions and only 1-2 water layers are present. In such small cells, higher dynamic variability was observed for the hydrophobic exposure and the backbone conformations as estimated from the Ramachandran regions. This effect is partly due to periodic image interactions and partly to the inability of one water layer to keep the peptide

folded. The first effect is technical, the second real, but the effects are not easily separated. The observation that small cell effects mainly affect hydrophobic exposure and not any other typical property studied by MD is however of interest by itself, regardless of the relative contribution of the two effects to this observation. In the large cells, including the cells of 10 Å, the hydrophobic effect plays out fully for this peptide, but we expect the critical requirement of water layers to increase with the solvent-exposed surface of the protein and thus its size. Small cells with only 1-2 layers of water around the peptide significantly impair the water-water interactions and consequently, the hydrophobic packing of the peptide.

Our conclusions are supported by studying also two other force fields, OPLS3e and Amber99sb-ILDN, showing that our division of properties into size-dependent and size-independent are not affected by the choice of the force field, even though the secondary structure propensity of the three force fields differs significantly, with Amber99sb-ILDN producing the least and OPLS3e the most helix.

To summarize, the most important findings of our work are that i) Aβ has very small barriers for structural transitions such that the peptide is sampled equally well in the explored part of the phase space by many short and fewer long simulations; we suggest that this is a defining hallmark of intrinsically disordered peptides; ii) most properties of interest are not cell size-dependent even for this very disordered, context-dependent peptide, and even when there is a possibility of periodic image interaction in small cells; this justifies the common use of 10-Å cells for simulating proteins; iii) for Aβ, the hydrophobic surface exposure is significantly larger in the very small cells (two-tailed t-test, 95% confidence) and is the only property to be affected by image interactions in our study; iv) we estimate that only 2-3 layers of water (as obtained in the 10-Å cells) are required to enforce the hydrophobic effect on this disordered peptide, and thus probably proteins more generally, possibly explaining why MD simulations using small simulation cells have been successful in the past.

SUPPORTING MATERIAL

The supporting information file contains all RMSD plots, RMSF plots and secondary structure plots for the simulations, including tables and figures analyzing the statistics of the ensembles. The data of secondary structure, radius of gyration, SASA, intra-peptide and peptide-water hydrogen bonds and Ramachandran plot analysis are provided as excel sheets in "**Supplementary Table S20**" for Charmm22*, "**Supplementary Table S31**" for OPLS3e and "**Supplementary Table S32**" for Amber99sb-ILDN potentials.

ACKNOWLEDGEMENTS

The Danish Council for Independent Research | Natural Sciences (DFF), grant case 7016-00079B, is acknowledged for supporting this work.

REFERENCES

[1] I.W. Hamley, Chem. Rev. **112**, 5147 (2012).

[2] J. Hardy, J. Alzheimer's Dis. **9**, 151 (2006).

[3] M.S. Wolfe, Biol. Chem. **393**, 899 (2012).

[4] J. Götz, A. Eckert, M. Matamales, L.M. Ittner, and X. Liu, Cell. Mol. Life Sci. **68**, 3359 (2011).

[5] J. Sevigny, P. Chiao, T. Bussière, P.H. Weinreb, L. Williams, M. Maier, R. Dunstan, S. Salloway, T. Chen, and Y. Ling, Nature **537**, 50 (2016).

[6] S. Salloway, R. Sperling, N.C. Fox, K. Blennow, W. Klunk, M. Raskind, M. Sabbagh, L.S. Honig, A.P. Porsteinsson, S. Ferris, M. Reichert, N. Ketter, B. Nejadnik, V. Guenzler, M. Miloslavsky, D. Wang, Y. Lu, J. Lull, I.C. Tudor, E. Liu, M. Grundman, E. Yuen, R. Black, and H.R. Brashear, N. Engl. J. Med. **370**, 322 (2014).

[7] M.D. Carter, G.A. Simms, and D.F. Weaver, Clin. Pharmacol. Ther. **88**, 475 (2010).

[8] J.-S. Choi, J.J. Braymer, R.P.R. Nanga, A. Ramamoorthy, and M.H. Lim, Proc. Natl. Acad. Sci. U. S. A. **107**, 21990 (2010).

[9] A.S. DeToma, S. Salamekh, A. Ramamoorthy, and M.H. Lim, Chem. Soc. Rev. **41**, 608 (2012).

[10] A. Ramamoorthy and M.H. Lim, Biophys. J. **105**, 287 (2013).

[11] M. Karplus and J.A. McCammon, Nat. Struct. Biol. **9**, 646 (2002).

[12] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, and T.E. Cheatham, Acc. Chem. Res. **33**, 889 (2000).

[13] J.P. Linge, M.A. Williams, C.A.E.M. Spronk, A.M.J.J. Bonvin, and M. Nilges, Proteins **50**, 496 (2003).

[14] M. Levitt and A. Warshel, Nature **253**, 694 (1975).

[15] J. Huang and A.D. MacKerell, Curr. Opin. Struct. Biol. **48**, 40 (2018).

[16] W.F. van Gunsteren and H.J.C. Berendsen, Angew. Chemie Int. Ed. English **29**, 992 (1990).

[17] D.A.C. Beck and V. Daggett, Methods **34**, 112 (2004).

[18] P.H. Hünenberger and J.A. McCammon, J. Chem. Phys. **110**, 1856 (1999).

[19] S.A. Adcock and J.A. McCammon, Chem. Rev. **106**, 1589 (2006).

[20] I.-C. Yeh and G. Hummer, J. Phys. Chem. B **108**, 15873 (2004).

[21] K. Takemura and A. Kitao, J. Phys. Chem. B **111**, 11870 (2007).

[22] D.P. Sellan, E.S. Landry, J.E. Turney, A.J.H. McGaughey, and C.H. Amon, Phys. Rev. B **81**, 214305 (2010).

[23] M. Sprik, J. Hutter, and M. Parrinello, J. Chem. Phys. **105**, 1142 (1996).

[24] O.N. de Souza and R.L. Ornstein, Biophys. J. **72**, 2395 (1997).

[25] J. Higo, H. Kono, N. Nakajima, H. Shirai, H. Nakamura, and A. Sarai, Chem. Phys. Lett. **306**, 395 (1999).

[26] K. El Hage, F. Hedin, P.K. Gupta, M. Meuwly, and M. Karplus, Elife **7**, e35560 (2018).

[27] V. Gapsys and B.L. de Groot, Elife **8**, e44718 (2019).

[28] K. El Hage, F. Hédin, P.K. Gupta, M. Meuwly, and M. Karplus, Elife **8**, e45318 (2019).

[29] M. Saito, J. Chem. Phys. **101**, 4055 (1994).

[30] P.J. Steinbach and B.R. Brooks, J. Comput. Chem. **15**, 667 (1994).

[31] R.J. Loncharich and B.R. Brooks, Proteins Struct. Funct. Bioinforma. **6**, 32 (1989).

[32] H. Schreiber and O. Steinhauser, Chem. Phys. **168**, 75 (1992).

[33] H. Schreiber and O. Steinhauser, J. Mol. Biol. **228**, 909 (1992).

[34] P.E. Smith and B.M. Pettitt, J. Chem. Phys. **95**, 8430 (1991).

[35] K.P. Jensen, J. Phys. Chem. B **112**, 1820 (2008).

[36] K. Lindorff-Larsen, P. Maragakis, S. Piana, M.P. Eastwood, R.O. Dror, and D.E. Shaw, PLoS One **7**, 1 (2012).

[37] J. Israelachvili and H. Wennerström, Nature **379**, 219 (1996).

[38] N.T. Southall, K.A. Dill, and A.D.J. Haymet, **106**, 521 (2002).

[39] C. Tsai, S.L. Lin, H.J. Wolfson, and R. Nussinov, Protein Sci. **6**, 53 (1997).

[40] J. Henriques, C. Cragnell, and M. Skepö, J. Chem. Theory Comput. **11**, 3420 (2015).

[41] I.M. Ilie and A. Caflisch, Chem. Rev. (2019).

[42] M. Chan-Yao-Chong, D. Durand, and T. Ha-Duong, J. Chem. Inf. Model. **59**, 1743 (2019).

[43] K.A. Ball, D.E. Wemmer, and T. Head-Gordon, J. Phys. Chem. B **118**, 6405 (2014).

[44] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. De Groot, H. Grubmüller, and A.D. MacKerell, Nat. Methods **14**, 71 (2016).

[45] J.-E. Shea and B. Urbanc, Curr. Top. Med. Chem. **12**, 2596 (2013).

[46] J. Khandogin and C.L. Brooks, Proc. Natl. Acad. Sci. U. S. A. **104**, 16880 (2007).

[47] R. Ramis, J. Ortega-Castro, R. Casasnovas, L. Mariño, B. Vilanova, M. Adrover, and J. Frau, J. Chem. Inf. Model. (2019).

[48] S. Vivekanandan, J.R. Brender, S.Y. Lee, and A. Ramamoorthy, Biochem. Biophys. Res. Commun. **411**, 312 (2011).

[49] M.F.M. Sciacca, S.A. Kotler, J.R. Brender, J. Chen, D.K. Lee, and A. Ramamoorthy, Biophys. J. **103**, 702 (2012).

[50] K.P. Kepp, Prog. Neurobiol. **143**, 36 (2016).

[51] M.K. Tiwari and K.P. Kepp, Alzheimer's Dement. **12**, 184 (2016).

[52] H. Lin, R. Bhatia, and R. Lal, FASEB J. **15**, 2433 (2001).

[53] S.K. Rhee, A.P. Quist, and R. Lal, J. Biol. Chem. **273**, 13379 (1998).

[54] K. Ono, M.M. Condron, and D.B. Teplow, Proc. Natl. Acad. Sci. U. S. A. **106**, 14745 (2009).

[55] C. Haass and D.J. Selkoe, Nat. Rev. Mol. Cell Biol. **8**, 101 (2007).

[56] D.J. Selkoe and J. Hardy, EMBO Mol. Med. **8**, 595 (2016).

[57] S.A. Kotler, P. Walsh, J.R. Brender, and A. Ramamoorthy, Chem. Soc. Rev. **43**, 8 (2014).

[58] A.K. Somavarapu and K.P. Kepp, ACS Chem. Neurosci. **6**, 1990 (2015).

[59] M.K. Tiwari and K.P. Kepp, Dalt. Trans. **44**, 2747 (2015).

[60] M.K. Tiwari and K.P. Kepp, J. Alzheimer's Dis. **47**, 215 (2015).

[61] I. Kuperstein, K. Broersen, I. Benilova, J. Rozenski, W. Jonckheere, M. Debulpaep, A. Vandersteen, I. Segers-Nolten, K. Van Der Werf, V. Subramaniam, D. Braeken, G. Callewaert, C. Bartic, R. D'Hooge, I.C. Martins, F. Rousseau, J. Schymkowitz, and B. De

Strooper, EMBO J. **29**, 3408 (2010).

[62] C.G. Glabe, Neurobiol. Aging **27**, 570 (2006).

[63] C. Poojari, A. Kukol, and B. Strodel, Biochim. Biophys. Acta - Biomembr. **1828**, 327 (2013).

[64] K.J. Korshavn, A. Bhunia, M.H. Lim, and A. Ramamoorthy, Chem. Commun. **52**, 882 (2016).

[65] R. Lal, H. Lin, and A.P. Quist, Biochim. Biophys. Acta **1768**, 1966 (2007).

[66] N. Arispe, H.B. Pollard, and E. Rojas, Proc. Natl. Acad. Sci. **93**, 1710 (1996).

[67] A. Baumketner, S.L. Bernstein, T. Wyttenbach, G.A.L. Bitan, D.B. Teplow, M.T. Bowers, and J. Shea, Filtration 420 (2006).

[68] S. Gnanakaran, R. Nussinov, and A.E. García, J. Am. Chem. Soc. **128**, 2158 (2006).

[69] D.D. Boehr, R. Nussinov, and P.E. Wright, Nat. Chem. Biol. **5**, 789 (2009).

[70] S.L. Bernstein, T. Wyttenbach, A. Baumketner, J.-E. Shea, G. Bitan, D.B. Teplow, and M.T. Bowers, J. Am. Chem. Soc. **127**, 2075 (2005).

[71] M. Yang and D.B. Teplow, J. Mol. Biol. **384**, 450 (2008).

[72] O. Coskuner-Weber and V. Uversky, Int. J. Mol. Sci. **19**, 336 (2018).

[73] Y.-Y. Wang, L. Li, T.-T. Chen, W.-Y. Chen, and Y.-C. Xu, Acta Pharmacol. Sin. **34**, 1243 (2013).

[74] D.J. Rosenman, C.R. Connors, W. Chen, C. Wang, and A.E. García, J. Mol. Biol. **425**, 3338 (2013).

[75] L.P. Savtchenko and D.A. Rusakov, Proc. Natl. Acad. Sci. **104**, 1823 (2007).

[76] J.D. Clements, Trends Neurosci. **19**, 163 (1996).

[77] K.P. Kepp, Coord. Chem. Rev. **351**, 127 (2017).

[78] B. Calabrese, G.M. Shaked, I. V Tabarean, J. Braga, E.H. Koo, and S. Halpain, Mol. Cell. Neurosci. **35**, 183 (2007).

[79] P.N. Lacor, M.C. Buniel, L. Chang, S.J. Fernandez, Y. Gong, K.L. Viola, M.P. Lambert, P.T. Velasco, E.H. Bigio, C.E. Finch, G.A. Krafft, and W.L. Klein, J. Neurosci. **24**, 10191 (2004).

[80] E. Abramov, I. Dolev, H. Fogel, G.D. Ciccotosto, E. Ruff, and I. Slutsky, Nat. Neurosci. **12**, 1567 (2009).

[81] Schrödinger and D. E. Shaw Research, *Desmond Molecular Dynamics System* (Schrödinger, LLC, New York, US, 2018).

[82] A.K. Somavarapu and K.P. Kepp, ChemPhysChem **16**, 3278 (2015).

[83] S. Piana, K. Lindorff-Larsen, and D.E. Shaw, Biophys. J. **100**, 47 (2011).

[84] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, J. Chem. Phys. **79**, 926 (1983).

[85] K.J. Bowers, F.D. Sacerdoti, J.K. Salmon, Y. Shan, D.E. Shaw, E. Chow, H. Xu, R.O. Dror, M.P. Eastwood, B.A. Gregersen, J.L. Klepeis, I. Kolossvary, and M.A. Moraes, in *Proc. 2006 ACM/IEEE Conf. Supercomput. - SC '06* (2006), p. 84.

[86] K.P. Kepp, J. Alzheimer's Dis. **55**, 447 (2017).

[87] G.J. Martyna, M.L. Klein, and M. Tuckerman, J. Chem. Phys. **97**, 2635 (1992).

[88] G.J. Martyna, D.J. Tobias, and M.L. Klein, J. Chem. Phys. **101**, 4177 (1994).

[89] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J.Y. Xiang, L. Wang, D. Lupyan, M.K. Dahlgren, J.L. Knight, J.W. Kaus, D.S. Cerutti, G. Krilov, W.L. Jorgensen, R. Abel, and R.A.

Friesner, J. Chem. Theory Comput. **12**, 281 (2016).

[90] K. Roos, C. Wu, W. Damm, M. Reboul, J.M. Stevenson, C. Lu, M.K. Dahlgren, S. Mondal, W. Chen, and L. Wang, J. Chem. Theory Comput. **15**, 1863 (2019).

[91] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, and D.E. Shaw, Proteins Struct. Funct. Bioinforma. **78**, 1950 (2010).

[92] D.M. Zuckerman, Annu. Rev. Biophys. **40**, 41 (2011).

[93] M.C. Zwier and L.T. Chong, Curr. Opin. Pharmacol. **10**, 745 (2010).

[94] J.R. Brender, E.L. Lee, K. Hartman, P.T. Wong, A. Ramamoorthy, D.G. Steel, and A. Gafni, Biophys. J. **100**, 685 (2011).

[95] R. Armen, D.O.V. Alonso, and V. Daggett, Protein Sci. **12**, 1145 (2003).

[96] S. Tomaselli, V. Esposito, P. Vangone, N.A.J. Van Nuland, A.M.J.J. Bonvin, R. Guerrini, T. Tancredi, P.A. Temussi, and D. Picone, ChemBioChem **7**, 257 (2006).

[97] D. Scheuner, C. Eckman, M. Jensen, X. Song, M. Citron, N. Suzuki, T.D. Bird, J. Hardy, M. Hutton, W. Kukull, E. Larson, E. Levy-Lahad, M. Viitanen, E. Peskind, P. Poorkaj, G. Schellenberg, R. Tanzi, W. Wasco, L. Lannfelt, D. Selkoe, and S. Younkin, Nat. Med. **2**, 864 (1996).

[98] M. Coles, W. Bicknell, A.A. Watson, D.P. Fairlie, and D.J. Craik, Biochemistry **37**, 11064 (1998).

[99] D.J. Rosenman, C. Wang, and A.E. Garcia, J. Phys. Chem. B **120**, 259 (2016).

[100] K.A. Beauchamp, Y.S. Lin, R. Das, and V.S. Pande, J. Chem. Theory Comput. **8**, 1409 (2012).

[101] E.A. Cino, W. Choy, and M. Karttunen, J. Chem. Theory Comput. **8**, 2725 (2012).

[102] M. Carballo-Pacheco and B. Strodel, Protein Sci. **26**, 174 (2017).

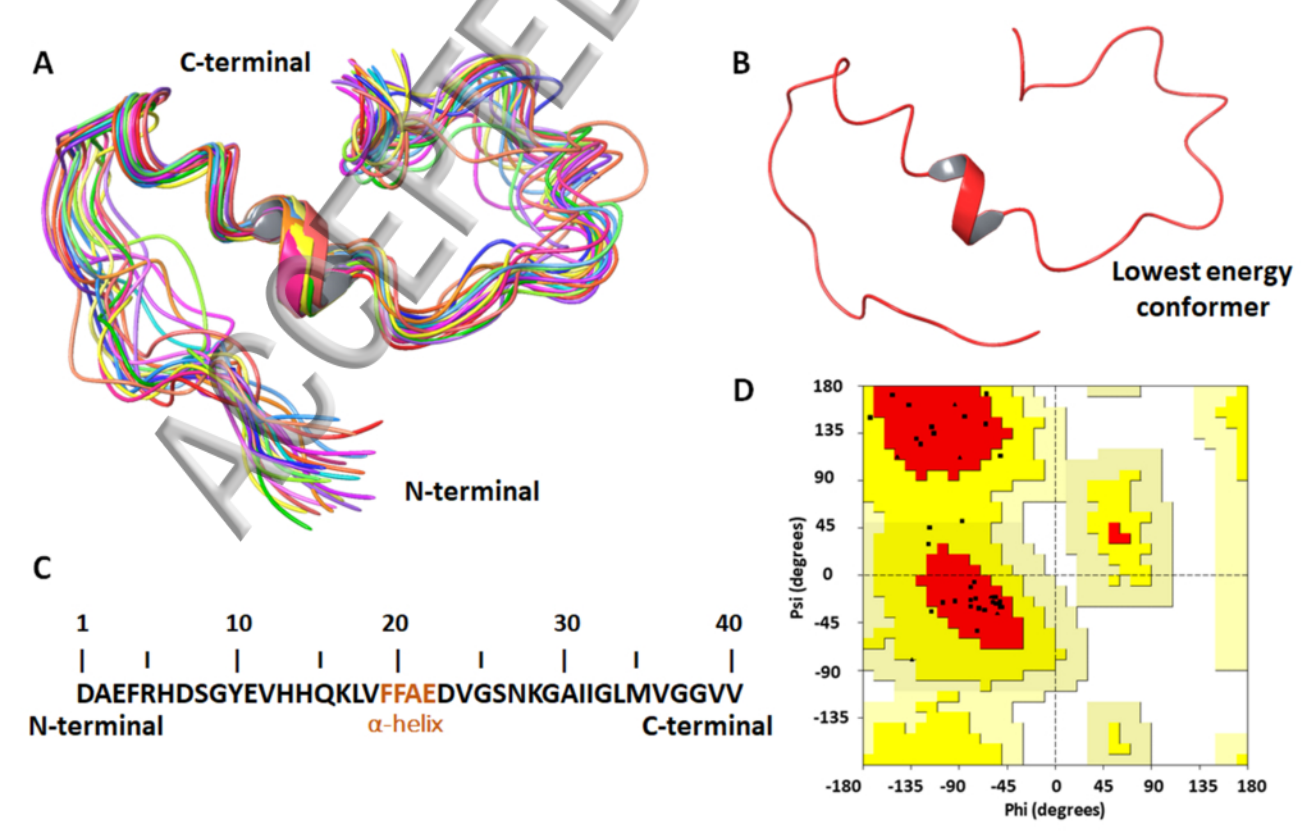
[103] O. Crescenzi, S. Tomaselli, R. Guerrini, S. Salvadori, A.M. D'Ursi, P.A. Temussi, and D. Picone, Eur. J. Biochem. **269**, 5642 (2002).

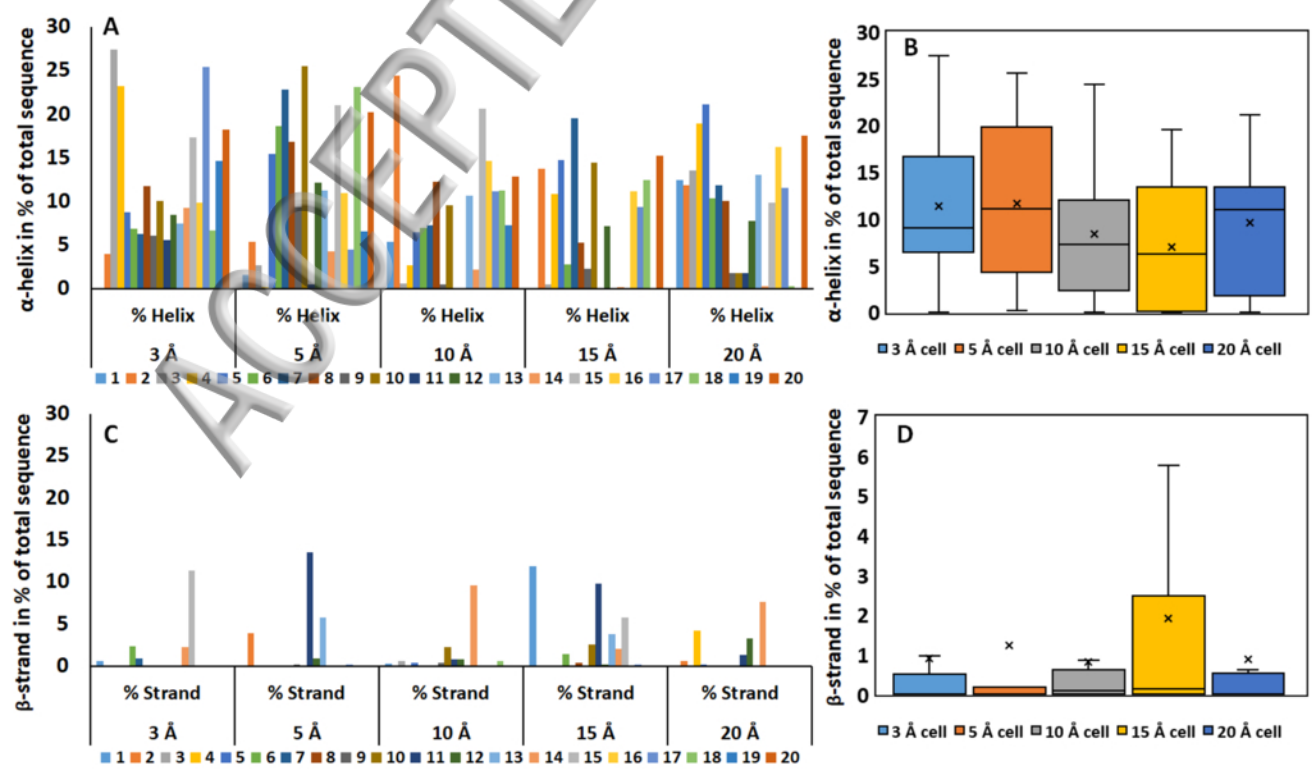
[104] H. Sticht, P. Bayer, D. Willbold, S. Dames, C. Hilbich, K. Beyreuther, R.W. Frank, and P. Rosch, Eur. J. Biochem. **233**, 293 (1995).

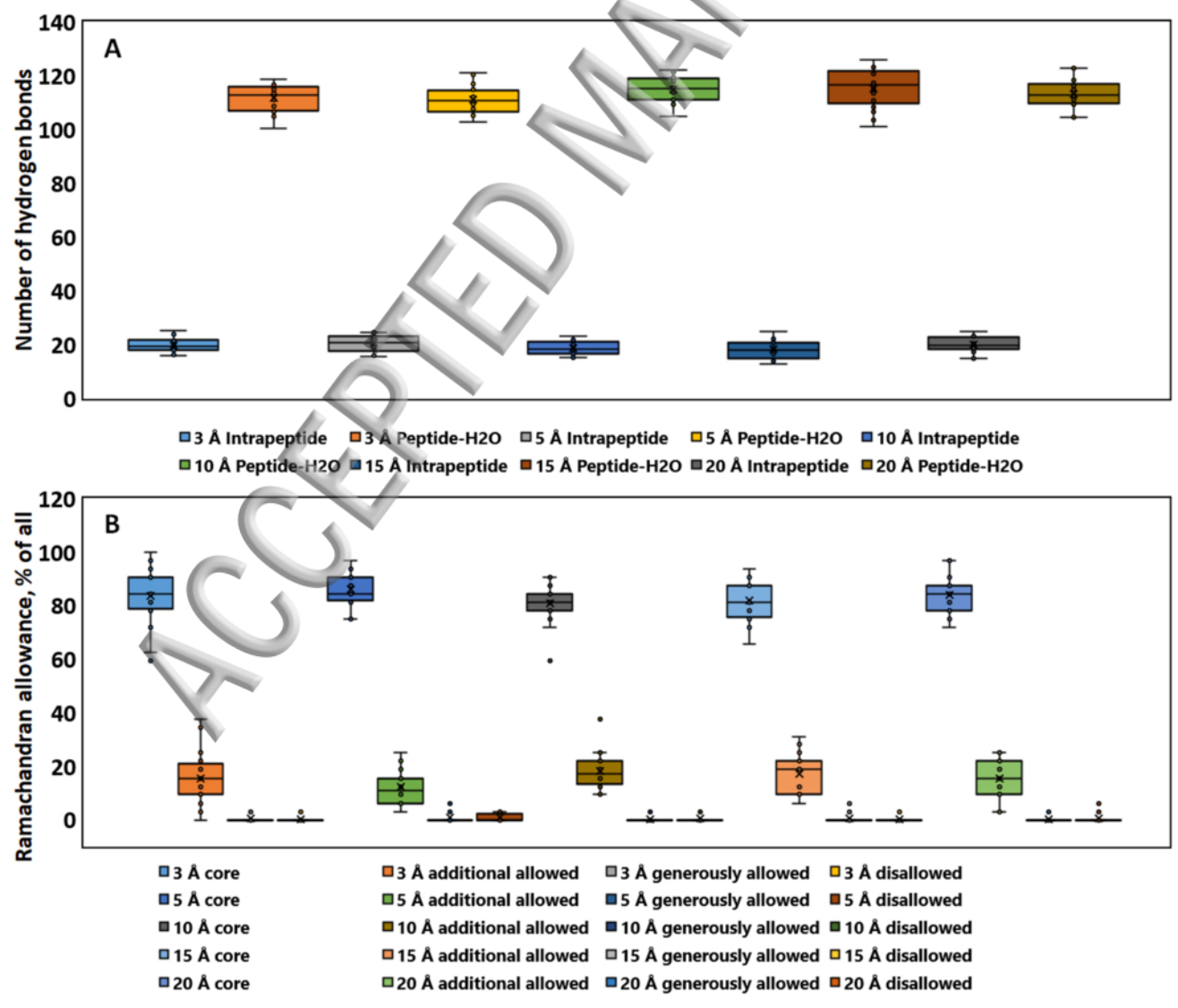
[105] L. Hou, H. Shao, Y. Zhang, H. Li, N.K. Menon, E.B. Neuhaus, J.M. Brewer, I.-J.L. Byeon, D.G. Ray, M.P. Vitek, T. Iwashita, R.A. Makula, A.B. Przybyla, and M.G. Zagorski, J. Am. Chem. Soc. **126**, 1992 (2004).

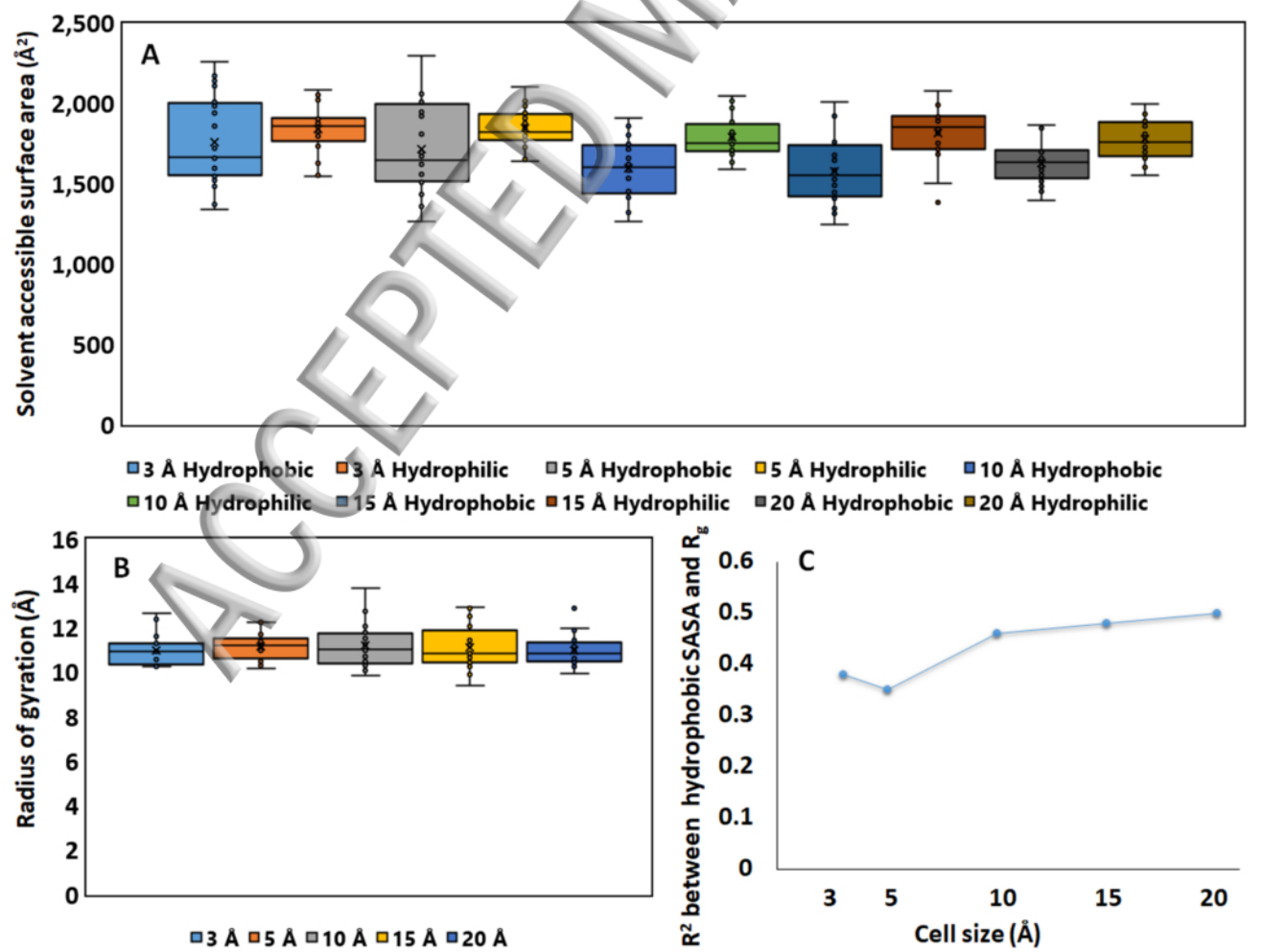
[106] Y. Yan, S.A. McCallum, and C. Wang, J. Am. Chem. Soc. **130**, 5394 (2008).

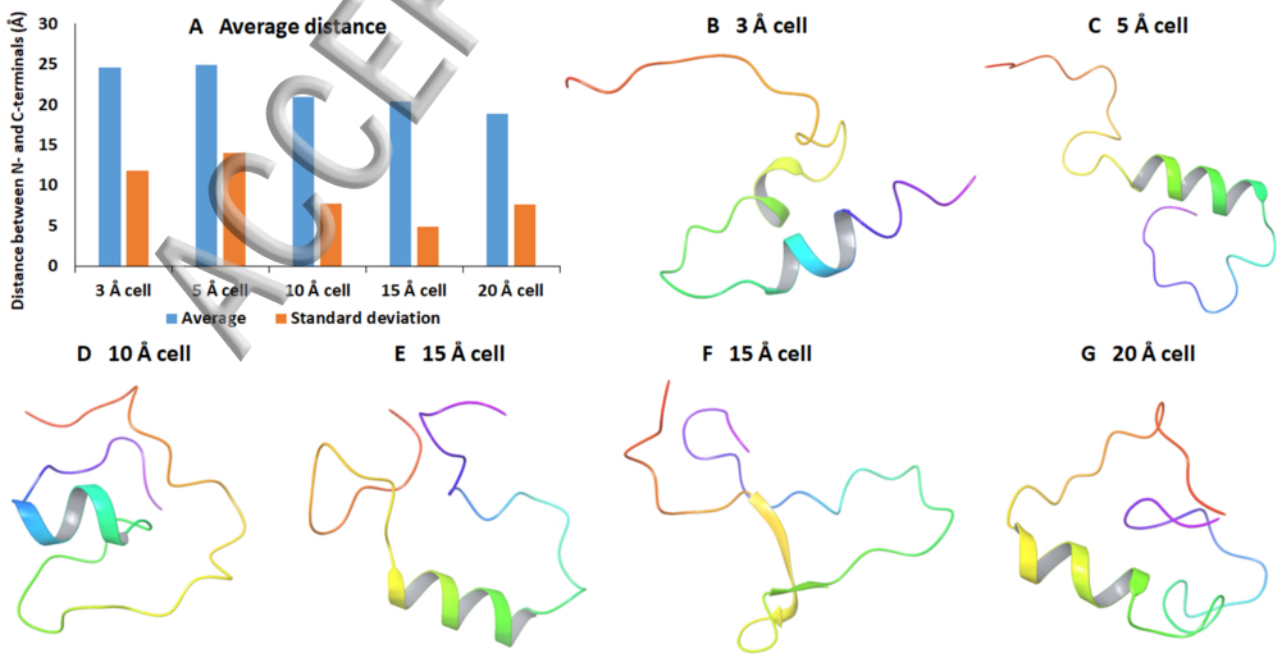
[107] A. Abelein, B. Bolognesi, C.M. Dobson, A. Graslund, and C. Lendel, Biochemistry **51**, 126 (2012).

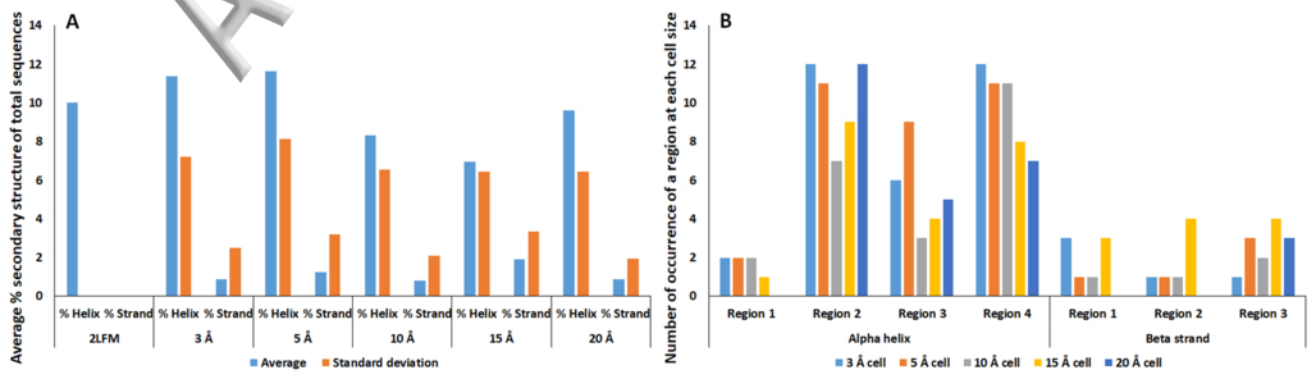

A
C-terminal
N-terminal
B
Lowest energy conformer
C
1
10
20
30
40
DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV
N-terminal
α-helix
C-terminal
D
Psi (degrees)
Phi (degrees)
180
135
90
45
0
-45
-90
-135
-180

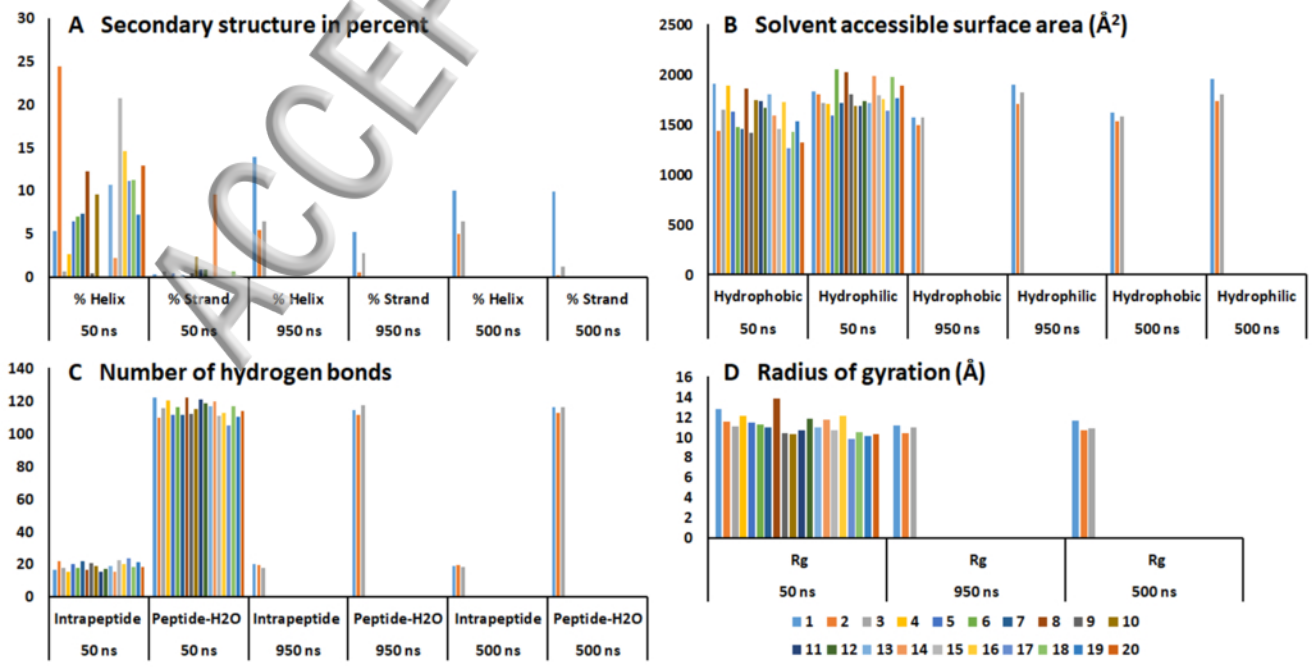

A
α-helix in % of total sequence
% Helix
3 Å
5 Å
10 Å
15 Å
20 Å
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
B
α-helix in % of total sequence
3 Å cell
5 Å cell
10 Å cell
15 Å cell
20 Å cell
C
β-strand in % of total sequence
% Strand
3 Å
5 Å
10 Å
15 Å
20 Å
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
D
β-strand in % of total sequence
3 Å cell
5 Å cell
10 Å cell
15 Å cell
20 Å cell


A
Number of hydrogen bonds
140
120
100
80
60
40
20
0
3 Å Intrapeptide
3 Å Peptide-H2O
5 Å Intrapeptide
5 Å Peptide-H2O
10 Å Intrapeptide
10 Å Peptide-H2O
15 Å Intrapeptide
15 Å Peptide-H2O
20 Å Intrapeptide
20 Å Peptide-H2O
B
Ramachandran allowance, % of all
120
100
80
60
40
20
0
3 Å core
3 Å additional allowed
3 Å generously allowed
3 Å disallowed
5 Å core
5 Å additional allowed
5 Å generously allowed
5 Å disallowed
10 Å core
10 Å additional allowed
10 Å generously allowed
10 Å disallowed
15 Å core
15 Å additional allowed
15 Å generously allowed
15 Å disallowed
20 Å core
20 Å additional allowed
20 Å generously allowed
20 Å disallowed


A
Solvent accessible surface area (Å²)
2,500
2,000
1,500
1,000
500
0
3 Å Hydrophobic
3 Å Hydrophilic
5 Å Hydrophobic
5 Å Hydrophilic
10 Å Hydrophobic
10 Å Hydrophilic
15 Å Hydrophobic
15 Å Hydrophilic
20 Å Hydrophobic
20 Å Hydrophilic
B
Radius of gyration (Å)
16
14
12
10
8
6
4
2
0
3 Å
5 Å
10 Å
15 Å
20 Å
C
R² between hydrophobic SASA and Rg
0.6
0.5
0.4
0.3
0.2
0.1
0
3
5
10
15
20
Cell size (Å)


A Average distance
Distance between N- and C-terminals (Å)
30
25
20
15
10
5
0
3 Å cell
5 Å cell
10 Å cell
15 Å cell
20 Å cell
Average
Standard deviation
B 3 Å cell
C 5 Å cell
D 10 Å cell
E 15 Å cell
F 15 Å cell
G 20 Å cell


A
Average % secondary structure of total sequences
14
12
10
8
6
4
2
0
% Helix
% Strand
2LFM
3 Å
5 Å
10 Å
15 Å
20 Å
Average
Standard deviation
B
Number of occurrence of a region at each cell size
Region 1
Region 2
Region 3
Region 4
Alpha helix
Beta strand
3 Å cell
5 Å cell
10 Å cell
15 Å cell
20 Å cell


A Secondary structure in percent
30
25
20
15
10
5
0
% Helix
50 ns
% Strand
50 ns
% Helix
950 ns
% Strand
950 ns
% Helix
500 ns
% Strand
500 ns
B Solvent accessible surface area (Å²)
2500
2000
1500
1000
500
0
Hydrophobic
50 ns
Hydrophilic
50 ns
Hydrophobic
950 ns
Hydrophilic
950 ns
Hydrophobic
500 ns
Hydrophilic
500 ns
C Number of hydrogen bonds
140
120
100
80
60
40
20
0
Intrapeptide
50 ns
Peptide-H2O
50 ns
Intrapeptide
950 ns
Peptide-H2O
950 ns
Intrapeptide
500 ns
Peptide-H2O
500 ns
D Radius of gyration (Å)
16
14
12
10
8
6
4
2
0
Rg
50 ns
Rg
950 ns
Rg
500 ns
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19 20